H

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/998,959	11/01/2001	Bishnu Saroop Atal	2000-0606	6485

26652 7590 08/26/2004

AT&T CORP.
P.O. BOX 4110
MIDDLETOWN, NJ 07748

EXAMINER	
ALBERTALLI, BRIAN LOUIS	
ART UNIT	PAPER NUMBER
2655	

DATE MAILED: 08/26/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/998,959	ATAL, BISHNU SAROOP	
	Examiner	**Art Unit**	
	Brian L Albertalli	2655	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1-36 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1-36 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on 01 November 2001 is/are: a)☐ accepted or b)☒ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date 11/1/01, 4/22/04.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____.

DETAILED ACTION

Specification

1. The disclosure is objected to because of the following informalities:

   a) On page 14, line 12, "Fig. 6" should be –Fig. 5--; and

   In line 18, "Fig. 5" should be –Fig. 6--.

   b) On page 17, line 5, "Fig. 6" should be –Fig. 8--.

   Appropriate correction is required.

Drawings

2. The drawings are objected to as failing to comply with 37 CFR 1.84(p)(5) because they include the following reference character(s) not mentioned in the description: 108. Corrected drawing sheets are required in reply to the Office action to avoid abandonment of the application. The objection to the drawings will not be held in abeyance.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1-2, 5, 25, 30-31, and 35-36 are rejected under 35 U.S.C. 103(a) as being unpatentable over Casey (U.S. Patent 6,321,200), in view of Smith et al. (*Template Adaptation in a Hypersphere Word Classifier*).

4. In regard to claim 1 and 36, Casey discloses a method for extracting features from a set of phonemes (column 4, lines 25-29), comprising:

Determining a phoneme vector (Fig. 1, band-pass signals 111) as a time-frequency representation of the class phoneme (step 110, column 3, lines 1-10);

Dividing the phoneme vector into phoneme segments (step 120, each bandpass signal 111 is windowed, column 3, lines 11-13);

Assigning each phoneme segment into a plurality of phoneme parameters (each window contains hundreds of parameters, see Fig. 2, 121, column 3, lines 13-14);

Expanding each phoneme segment and plurality of phoneme parameters into an expanded stored phoneme vector with expanded vector parameters (spectral features of observation matrix 121 are expressed as vectors, column 3, lines 50-53); and

Transforming the stored-phoneme vector (observation matrix 121) into an orthogonal form using singular-value decomposition (step 130, column 3, lines 16-31).

Casey further discloses the method is used for extracting features form analog acoustic signals (column 1, lines 66-67) and must, inherently, convert the analog acoustic signal to a digital signal to process the signal.

Casey does not disclose that the extracted features are used to train class phonemes or recognize class phonemes, wherein the recognition was determined by:

Determining a first distance associated with the orthogonal form of the expanded received-signal vector and a second distance associated respectively with each orthogonal form of the expanded stored-phoneme vectors; and

Recognizing the received phoneme according to a comparison of the first distance with the second distance.

Smith et al. discloses a method of speech recognition that utilizes a hyperspheres as templates for word recognition. The method includes a training phase (see page 565, Template Generation section) and a recognition phase (Matching section).

The recognition phase of Smith's method comprises:

Determining a first distance associated with the received-signal vector and a second distance associated respectively with the expanded stored-phoneme vectors (template of the representing the input is compared with each of the templates in the vocabulary); and

Recognizing the received phoneme according to a comparison of the first distance with the second distance (the vocabulary template with the best score is used as the recognition result, page 565, second column, Matching section).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Casey so that the method of extracting features, as disclosed by Casey, was used to create the template patterns used in speech pattern

training and recognition, as disclosed by Smith et al., so that a highly accurate representation of speech could be used in the speech recognizer, thereby increasing the chances of correct recognition results.

5. In regard to claim 2, Casey discloses transforming the stored-phoneme vector (observation matrix 121) into an orthogonal form using singular-value decomposition (step 130, column 3, lines 16-31).

Casey does not disclose that transforming the expanded received-signal vector into an orthogonal form using singular-value decomposition conforms the stored-phoneme vector and the expanded received-signal vector into a hypersphere having a center and a radius.

Smith et al. discloses that a stored-phoneme vector and a received-signal vector (templates) are n-dimensional hyperspheres (page 565, Description of the Recognizer section), which applies also to their orthogonal forms. A hypersphere must, by definition, have a center and a radius.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Casey to conform the stored phoneme vector into a hypersphere, since representing stored-phoneme vectors and received-phoneme vectors as hyperspheres simplifies the process of adapting the stored-phoneme vectors (templates) for better recognition results, as taught by Smith et al. (pages 565-566, Adaptive Training section).

6. In regard to claim 5, Casey discloses the orthogonal form of the expanded stored-phoneme vector and the expanded received-signal vector each have at least approximately 100 dimensions.

Casey discloses that 40 to 50 spectral parameters are included in each observation matrix, which includes hundreds of samples (column 3, lines 4-7, and lines 13-14). This would correspond to at least a 4000 dimensional (100 samples * 40 frequency bands) matrix. Although Casey discloses that the dimensionality of the observation matrix is reduced by the SVD, in Fig. 4, a graph of the vector representing spectral features of an input signal clearly has at least 500 frequency bins, which would correspond to an approximately 500 dimensional expanded received-signal vector. The stored-phoneme vector would be created in the same manner as the expanded received-signal vector.

7. In regard to claim 25 and 35, Casey discloses:

Receiving a received phoneme (audio mixture 101, column 3, lines 1-4); and

Recognizing the received phoneme according to a comparison of the received phoneme to each of the stored phonemes (column 4, lines 25-29).

Casey does not disclose:

Converting the received phoneme to n-dimensional space; and

Comparing the received phoneme to each of the stored phonemes in n-dimensional space.

Smith et al. discloses converting the received phoneme to n-dimensional space (template representing the input, page 565, Description of the Recognizer section); and

Comparing the received phoneme to each of the stored phonemes in n-dimensional space (an input template is matched against each template in the vocabulary, page 565, Matching section).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Casey so the comparison between a sampled phoneme and the converted plurality of phonemes was a measurement of the distance between the sampled phoneme and the converted plurality of phonemes in n-dimensional space, since by representing a phoneme in n-dimensional space, adapting the converted plurality of phonemes (templates) only requires adding the rejected input patterns to the closest template, as taught by Smith et al. (page 566, Template Subtraction section, lines 1-6).

8. In regard to claim 30, Casey discloses a system comprising:

A recording element that receives a phoneme (analog acoustic signals are input, column 1, lines 66-67, the extracted features of which are used for phoneme recognition, column 4, lines 25-29).

Casey does not disclose a computer that converts the received phoneme into n-dimensional space, wherein the computer compares in the n-dimensional space the received phoneme with each phoneme in the database of stored phonemes.

Smith et al. discloses a computer that converts the received phoneme to n-dimensional space (template representing the input, page 565, Description of the Recognizer section) wherein the computer compares in the n-dimensional space the received phoneme with each phoneme in the database of stored phonemes (an input template is matched against each template in the vocabulary, page 565, Matching section).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Casey so the comparison between a sampled phoneme and the converted plurality of phonemes was a measurement of the distance between the sampled phoneme and the converted plurality of phonemes in n-dimensional space, since by representing a phoneme in n-dimensional space, adapting the converted plurality of phonemes (templates) only requires adding the rejected input patterns to the closest template, as taught by Smith et al. (page 566, Template Subtraction section, lines 1-6).

9. In regard to claim 31, the combination of Casey and Smith et al., as applied to claim 30, above, discloses in Smith et al. that the computer recognizes the received phoneme (input) using the comparison in the n-dimensional space of the received phoneme from the database of stored phonemes (templates, page 565, Matching section).

10. Claims 3-4, 6-7, 16-22, 25-29, and 32-34 are rejected under 35 U.S.C. 103(a) as being unpatentable over Casey, in view of Smith et al., and further in view of Cooper (*The Hypersphere in Pattern Recognition*).

In regard to claim 3, neither Casey nor Smith et al. disclose that determining a distance comprises comparing a distance from the center of the hypersphere of the orthogonal form of the expanded received-signal vector with a distance from the center of the hypersphere for each orthogonal form of the expanded stored-phoneme vector.

Cooper discloses that when using a hypersphere as a classification boundary, the distance of an unknown vector x from the hypersphere is determined by calculating the distance of the unknown vector x from the center of the hypersphere (page 326, section II, lines 1-10, equation 2).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey and Smith et al. so the distance between a stored-signal phoneme vector orthogonal form and a received-signal phoneme vector orthogonal form was measured comparing the center of the stored-signal phoneme vector and each received-signal phoneme vector, because comparing a threshold with the distance between an unknown and a fixed point, such as the center of the hypersphere, can be an excellent approximation decision boundary, as taught by Cooper (page 325 lines 21-26).

11. In regard to claim 4, the combination of Casey, Smith et al. and Cooper, as applied to claim 3, above, discloses in Cooper that the hypersphere can be

used for multiple category classification (page 338-339, Multiple Category section).

Neither Casey, Smith et al., nor Cooper disclose that the m-shortest differences between the center of the hypersphere of the received-signal vector and the center of the hypersphere for each orthogonal form of the expanded stored-phoneme vectors are recognized as most likely to be associated with the received phoneme.

Official notice is taken that it is notoriously well known in the art to consider several of the best choices as a possible recognition result. In a system which used a distance measurement to determine the similarity between an input phoneme and a stored phoneme, this would correspond to the m-shortest differences between the input phoneme and the stored phoneme.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey, Smith et al., and Cooper, so the m-shortest differences (which would be the m-best recognition choices) would be recognized as most likely to be associated with the received phoneme, so that if the shortest distance was determined to be an incorrect recognition result during later processing (such as during word recognition or phrase recognition) the next shortest distance result could be used.

12. In regard to claims 6 and 7, neither Casey nor Smith et al. disclose removing the mean value of a stored phoneme or a received phoneme vector.

Cooper discloses that for two distributions having the same mean, the hyperspheres corresponding with those distributions will have the same center. Therefore, the comparison between two distributions with the same mean only requires a calculation of the radius of the hypersphere (page 329, section C, lines 1-3 and page 330, lines 1-4).

It would have been obvious to one of ordinary skill in the art at the time of invention to remove the means of the stored phoneme vectors and the received phoneme vectors so the comparison of the received vectors to the stored vectors would only require a calculation of the radius of each.

13. In regard to claim 16, Casey discloses a method for recognizing speech patterns, the pattern comprising:

Sampling speech patterns to obtain at least one sampled phoneme (column 4, lines 25-29).

Casey does not disclose:

Converting each stored phoneme into n-dimensional space having a center;

Converting each of the at least one sampled phonemes into the n-dimensional space; and

Comparing a distance from the center of the n-dimensional space to the sampled phoneme with a distance from the center of the n-dimensional space to each of the phonemes of the converted plurality of phonemes.

Smith et al. discloses:

Converting each stored phoneme (template) into n-dimensional space having a center (page 565, Description of the Recognizer section);

Converting each of the at least one sampled phonemes into the n-dimensional space (template representing the input, page 565, Matching section); and

Comparing a distance of the n-dimensional space of a sampled phoneme (input) to each of the phonemes of the converted plurality of phonemes (templates, page 565, Matching section).

Smith et al. does not disclose that the distance is a distance from the center of the n-dimensional space of the sampled phoneme to the center of each of the phonemes of the converted plurality of phonemes.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Casey so the comparison between a sampled phoneme and the converted plurality of phonemes was a measurement of the distance between the sampled phoneme and the converted plurality of phonemes in n-dimensional space, since by representing a phoneme in n-dimensional space, adapting the converted plurality of phonemes (templates) only requires adding the rejected input patterns to the closest template, as taught by Smith et al. (page 566, Template Subtraction section, lines 1-6).

Cooper discloses that when using a hypersphere as a classification boundary, the distance of an unknown vector x from the hypersphere is determined by calculating the distance of the unknown vector x from the center of the hypersphere (page 326, section II, lines 1-10, equation 2).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey and Smith et al. so the distance between a stored-signal phoneme vector and a received-signal phoneme vector was measured comparing the center of the stored-signal phoneme vector and each received-signal phoneme vector, because comparing a threshold with the distance between an unknown and a fixed point, such as the center of the hypersphere, can be an excellent approximation boundary, as taught by Cooper (page 325 lines 21-26).

14. In regard to claim 17, Casey discloses transforming the stored-phoneme matrix (observation matrix 121) into an orthogonal form using singular-value decomposition (step 130, column 3, lines 16-31).

Casey does not disclose that the orthogonal form of the stored phoneme is then used to convert the stored phoneme into n-dimensional space.

Smith et al. discloses converting an input pattern into n-dimensional space (page 565, Description of the Recognizer section).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Casey to convert the orthogonal form of the stored phoneme vector into n-dimensional space, since by representing a phoneme in n-dimensional space, adapting the converted plurality of phonemes (templates) only requires adding the rejected input patterns to the closest template, as taught by Smith et al. (page 566, Template Subtraction section, lines 1-6).

15. In regard to claim 18, Casey discloses storing the converted phonemes before sampling speech patterns (extracted features of the sampled speech pattern is compared to a set of a-priori classes, column 4, lines 25-29).

16. In regard to claim 19, Casey discloses n equals at least 100.

Casey discloses that 40 to 50 spectral parameters are included in each observation matrix, which includes hundreds of samples (column 3, lines 4-7, and lines 13-14). This would correspond to at least a 4000 dimensional (100 samples * 40 frequency bands) matrix. Although Casey discloses that the dimensionality of the observation matrix is reduced by the SVD, in Fig. 4, a graph of the vector representing spectral features of an input signal clearly has at least 500 frequency bins, which would correspond to an approximately 500 dimensional expanded received signal vector.

17. In regard to claim 20, the combination of Casey, Smith et al., and Cooper, as applied to claim 16, above, discloses in Smith et al. that comparing the distance from the center of the n-dimensional space to the sampled phoneme with the distance from the center of the n-dimensional space to each of the converted phonemes further comprises: determining a difference between the distance from the center of the n-dimensional space to the sampled phoneme with the distance from the center of the n-dimensional space to each of the converted phonemes (the comparison between an input template and a template

in the recognizers vocabulary is the difference between the two templates, page 565, Matching section).

18. In regard to claim 21, the combination of Casey, Smith et al., and Cooper, as applied to claim 16, above, discloses in Smith et al. recognizing the sampled phoneme as the stored phoneme associated with the smallest difference between the distance from the center of the n-dimensional space to the sampled phoneme with the distance from the center of the n-dimensional space to each of the converted phonemes (the input is classified by the vocabulary template with the lowest score, page 565, Matching section).

19. In regard to claim 22, the combination of Casey, Smith et al., and Cooper, as applied to claim 16, above, discloses in Smith et al. that the n-dimensional space is hyperspherical (page 565, Description of the Recognizer section).

20. In regard to claim 26, the combination of Casey and Smith et al., as applied to claim 25, above, discloses, in Smith et al. determining a first distance associated with the received phoneme and a second distance associated with the stored phonemes (template of the representing the input is compared with each of the templates in the vocabulary).

Neither Casey nor Smith et al. disclose that the distance is a measurement from the center of the hypersphere.

Cooper discloses that when using a hypersphere as a classification boundary, the distance of an unknown vector x from the hypersphere is determined by calculating the distance of the unknown vector x from the center of the hypersphere (page 326, section II, lines 1-10, equation 2).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey and Smith et al. so the distance between a stored-signal phoneme vector and a received-signal phoneme vector was measured comparing the center of the stored-signal phoneme vector and each received-signal phoneme vector, because comparing a threshold with the distance between an unknown and a fixed point, such as the center of the hypersphere, can be an excellent approximation boundary, as taught by Cooper (page 325 lines 21-26).

21. In regard to claim 27, Casey discloses n equals at least 100.

Casey discloses that 40 to 50 spectral parameters are included in each observation matrix, which includes hundreds of samples (column 3, lines 4-7, and lines 13-14). This would correspond to at least a 4000 dimensional (100 samples * 40 frequency bands) matrix. Although Casey discloses that the dimensionality of the observation matrix is reduced by the SVD, in Fig. 4, a graph of the vector representing spectral features of an input signal clearly has at least 500 frequency bins, which would correspond to an approximately 500 dimensional expanded received signal vector.

22. In regard to claim 28, the combination of Casey, Smith et al., and Cooper, as applied to claim 25, above, discloses in Smith et al. determining the difference between the first distance and the second distance for each stored phoneme (the comparison between an input template and a template in the recognizers vocabulary is the difference between the two templates, page 565, Matching section).

23. In regard to claim 29, the combination of Casey, Smith et al., and Cooper, as applied to claim 25, above, discloses in Smith et al. recognizing the received phoneme according to the stored phoneme associated with the smallest difference between the first distance and the second distance (the input is classified by the vocabulary template with the lowest score, page 565, Matching section).

24. In regard to claim 32, the combination of Casey and Smith et al., as applied to claim 30, above, discloses in Smith et al. determining a distance in n-dimensional space to a first point associated with the received phoneme with a second distance associated with each respective stored phoneme (input is matched against each of the templates in the vocabulary, page 565, Matching section).

Neither Casey nor Smith et al. disclose that determining a distance comprises comparing a first distance from a center of the n-dimensional space to a first point associated with the received phoneme with a second distance from

the center of the n-dimensional space to a second point associated with each respective stored phoneme from the database of stored phonemes.

Cooper discloses that when using a hypersphere as a classification boundary, the distance of an unknown vector x from the hypersphere is determined by calculating the distance of the unknown vector x from the center of the hypersphere (page 326, section II, lines 1-10, equation 2).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey and Smith et al. so the distance between a stored-signal phoneme vector and a received-signal phoneme vector was measured comparing the center of the stored-signal phoneme vector and each received-signal phoneme vector, because comparing a threshold with the distance between an unknown and a fixed point, such as the center of the hypersphere, can be an excellent approximation boundary, as taught by Cooper (page 325 lines 21-26).

25. In regard to claim 33, the combination of Casey, Smith et al., and Cooper, as applied to claim 30, above, discloses in Smith et al. determining the difference between the first distance and the second distance for each stored phoneme (the comparison between an input template and a template in the recognizers vocabulary is the difference between the two templates, page 565, Matching section).

26. In regard to claim 34, the combination of Casey, Smith et al., and Cooper, as applied to claim 30, above, discloses in Smith et al. recognizing the received phoneme as associated with a stored phoneme corresponding to a shortest distance between the first distance and the second distance (the input is classified by the vocabulary template with the lowest distance score, page 565, Matching section).

27. Claims 8-15 and 23-24 are rejected under 35 U.S.C. 103(a) as being unpatentable over Casey, in view of Smith et al., in view of Cooper, and further in view of Ostendorf (*A Stochastic Segment Model for Phoneme-Based Continuous Speech Recognition*).

28. In regard to claim 8, Casey discloses recognizing phonemes (column 4, lines 25-29).

Casey, Smith et al. and Cooper do not disclose that the phoneme vector determined as a time-frequency representation of the class phoneme is a representation of approximately 125 msec.

Ostendorf discloses a phoneme recognition method that determines a time-frequency representation of a class phoneme that is approximately 125 msec (frames of speech are analyzed every 10 msec, page 1864, second column, third paragraph; and the average number of samples per phoneme is about 10, page 1859, second column, second paragraph, lines 5-10. This corresponds to a time-frequency representation of 100 msec.)

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey, Smith et al., and Cooper to determine a time-frequency representation of a class phoneme that was approximately 125 msec long, since this is a good approximation of the length of a phoneme, therefore all of the information necessary for recognition of that phoneme would be included in the 125 msec window.

29. In regard to claim 9, Casey discloses the phoneme vector is divided into approximately 25 msec phoneme segments (20 msec segments, column 3, lines 11-13).

30. In regard to claim 10, Casey discloses each phoneme segment is assigned approximately 32 phoneme parameters (each 20 msec phoneme segment is assigned 40-50 parameters, column 3, lines 4-7).

31. In regard to claim 11, the combination of Casey, Smith et al., Cooper, and Ostendorf would produce and expanded-stored phoneme with approximately 160 parameters.

An approximately 125 msec time-frequency representation (100 msec) as disclosed by Ostendorf, as applied to claim 8, above, would be windowed every 20 msec, as disclosed by Casey (column 3, lines 11-13), each window having about 40 spectral parameters, as disclosed by Casey (column 3, lines 4-7). This would result in a vector of 200 parameters.

32. In regard to claim 12, Casey, Smith et al. and Cooper do not disclose that the received-signal vector determined as a time-frequency representation of the class phoneme is a representation of approximately 125 msec.

Ostendorf discloses a phoneme recognition method that determines a time-frequency representation of a received-signal vector that is approximately 125 msec (frames of speech are analyzed every 10 msec, page 1864, second column, third paragraph; and the average number of samples per phoneme is about 10, page 1859, second column, second paragraph, lines 5-10. This corresponds to a time-frequency representation of 100 msec.)

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey, Smith et al., and Cooper to determine a time-frequency representation of a received-signal vector that was approximately 125 msec long, since this is a good approximation of the length of a phoneme, therefore all of the information necessary for recognition of that phoneme would be included in the 125 msec window.

33. In regard to claim 13, Casey discloses the received phoneme vector is divided into approximately 25 msec phoneme segments (20 msec segments, column 3, lines 11-13).

34. In regard to claim 14, Casey discloses each phoneme segment is assigned approximately 32 phoneme parameters (each 20 msec phoneme segment is assigned 40-50 parameters, column 3, lines 4-7).

35. In regard to claim 15, the combination of Casey, Smith et al., Cooper, and Ostendorf, as applied to claim 12, above, would produce an expanded received-signal vector with approximately 160 parameters.

An approximately 125 msec time-frequency representation (100 msec) as disclosed by Ostendorf, as applied to claim 8, above, would be windowed every 20 msec, as disclosed by Casey (column 3, lines 11-13), each window having about 40 spectral parameters, as disclosed by Casey (column 3, lines 4-7). This would result in a vector of 200 parameters.

36. In regard to claim 23, the combination of Casey, Smith et al., and Cooper do not disclose that a stored phoneme vector would have 160 parameters.

Ostendorf discloses a phoneme recognition method that determines a time-frequency representation of a stored phoneme vector that is approximately 125 msec (frames of speech are analyzed every 10 msec, page 1864, second column, third paragraph; and the average number of samples per phoneme is about 10, page 1859, second column, second paragraph, lines 5-10. This corresponds to a time-frequency representation of 100 msec.)

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey, Smith et al., and Cooper to

determine a time-frequency representation of a received-signal vector that was approximately 125 msec long, since this is a good approximation of the length of a phoneme, therefore all of the information necessary for recognition of that phoneme would be included in the 125 msec window.

The combination of Casey, Smith et al., Cooper, and Ostendorf, therefore, would produce an expanded received-signal vector with approximately 160 parameters.

An approximately 125 msec time-frequency representation (100 msec) as disclosed by Ostendorf, as applied to claim 8, above, would be windowed every 20 msec, as disclosed by Casey (column 3, lines 11-13), each window having about 40 spectral parameters, as disclosed by Casey (column 3, lines 4-7). This would result in a vector of 200 parameters. Smith et al. discloses creating a template for each stored word (page 565, Template Generation section).

Furthermore, Smith et al. discloses transforming the stored vector into the n-dimensional space wherein the probability density of the stored phonemes in the n-dimensional space is approximately spherical (pattern templates are represented as n-dimensional hyperspheres, page 565, Description of the Recognizer section).

37. In regard to claim 24, the combination of Casey, Smith et al., and Cooper do not disclose that a sampled phoneme would have 160 parameters.

Ostendorf discloses a phoneme recognition method that determines a time-frequency representation of a sampled phoneme vector that is

approximately 125 msec (frames of speech are analyzed every 10 msec, page 1864, second column, third paragraph; and the average number of samples per phoneme is about 10, page 1859, second column, second paragraph, lines 5-10. This corresponds to a time-frequency representation of 100 msec.)

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey, Smith et al., and Cooper to determine a time-frequency representation of a sampled phoneme vector that was approximately 125 msec long, since this is a good approximation of the length of a phoneme, therefore all of the information necessary for recognition of that phoneme would be included in the 125 msec window.

The combination of Casey, Smith et al., Cooper, and Ostendorf, therefore, would produce an expanded sampled phoneme vector with approximately 160 parameters.

An approximately 125 msec time-frequency representation (100 msec) as disclosed by Ostendorf, as applied to claim 8, above, would be windowed every 20 msec, as disclosed by Casey (column 3, lines 11-13), each window having about 40 spectral parameters, as disclosed by Casey (column 3, lines 4-7). This would result in a vector of 200 parameters. Smith et al. discloses creating a template for each stored word (page 565, Template Generation section).

Furthermore, Smith et al. discloses transforming the stored vector into the n-dimensional space wherein the probability density of the stored phonemes in the n-dimensional space is approximately spherical (pattern templates are

represented as n-dimensional hyperspheres, page 565, Description of the Recognizer section).

Conclusion

38. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Campbell et al. (U.S. Patent 5,946,653) discloses a method for speech recognition that performs a polynomial expansion on a received feature vector. Beigi et al. (U.S. Patent 6,246,982) discloses a method for measuring the distance between collections of probability distributions in n-dimensional space. Aldersberg (U.S. Patent 4,907,276) discloses a recognition system that searches for a match of an input vector within a hypersphere in n-dimensional space.

39. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Brian L Albertalli whose telephone number is (703) 305-1817. The examiner can normally be reached on Monday - Friday, 8:30 AM - 5:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Smits can be reached on (703) 305-3011. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

BLA 8/16/04

TĀLIVALDIS IVARS ŠMITS
PRIMARY EXAMINER